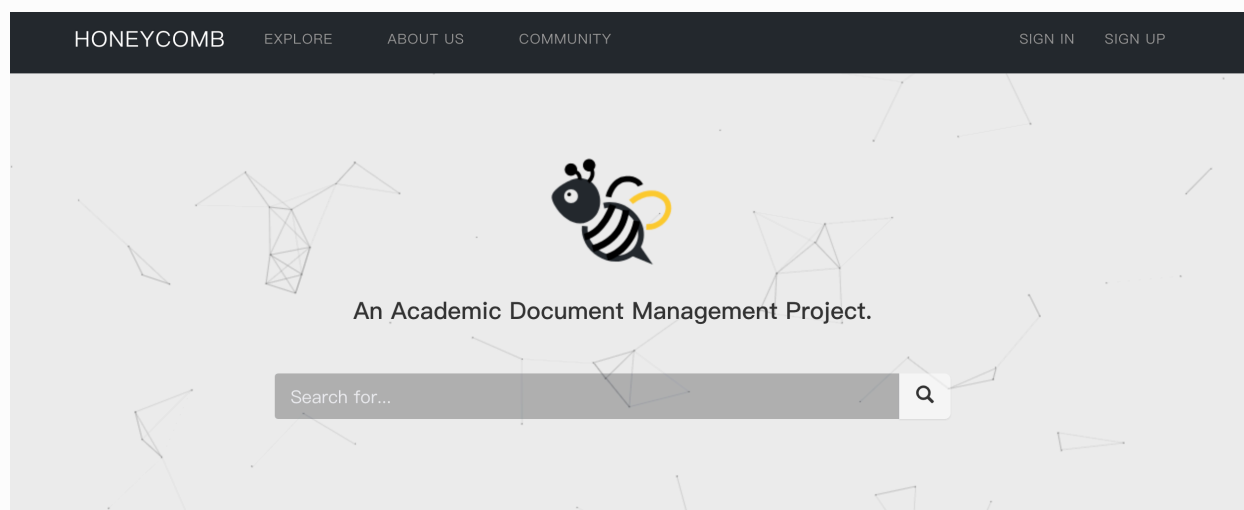


# 第二周工作报告

## 已完成工作


在第二周的时间区段内，我们小组已经完成的主要工作总结如下：

- 完整地达成了第一次迭代计划既定的任务内容：前端页面的搭建，数据库的定义，登录和注册功能，基础的交互界面等；
- 完成了对于论文搜索时需要的匹配查找算法的编写，算法的基础是文本和检索中常见的‘TF-IDF’算法；
- 完成了基于SSH框架的后端功能模块的基本编写，结合已经拥有的数据库数据，可以对包括用户登录、用户账号注册、个人头像上传更改等功能的支持，以及对于论文数据的匹配查找和排序等。其中，对于结构化数据的存储使用了Mysql数据库，对于包括用户头像在那的blob数据存储，结合了Mongodb数据库；
- 完成了对于基本的论文数据库内容的获取和填充，基于自主编写的爬虫程序，目前已经获得了总共16W条的论文数据，基础的数据信息有论文的标题、被引用量和部分论文的相互引用关系等。并且下一步的数据填充工作已经开启，主要是获取论文的完整作者目录和abstract等信息；
- 完成了数据库可视化的初步工作，创建了基于Vue.js和附加模块的前端数据可视化界面，对于需要的Pagerank算法进行了调研和实现；
- 完成了对于移动端上的网站页面显示的优化工作；
- 在小组购买的云服务器上搭建了工程所需要的环境并且上传了工程的雏形版本，在云服务器上得到了预想的运行状态，并且将服务器与开放的域名进行了绑定。目前，我们的web项目的雏形版本已经可以通过访问[www.combsite.site](http://www.combsite.site)网站进行浏览体验。对部分界面的展示如下：



HONEYCOMBEXPLOREABOUT USCOMMUNITY

Login



An Academic Document Management Project.

Search for...

User Name

User

Password

Password


☐ Remember Me

[No Account?](#)

Login

HONEYCOMBEXPLOREABOUT USCOMMUNITY

Register



An Academic Document Management Project.

Search for...

User Name

User

Email

Email

Password

Password

Confirm Your Password

Password Again

HONEYCOMBEXPLOREABOUT USCOMMUNITY

CJK >


cjk

Oops, empty here!

Edit Bio


Edit Profile

NetsCollectsTrends




**Intel® Software Guard Extensions**

This Intel technology is for application developers who are seeking to protect select code and data from disclosure or modification. Intel® SGX makes such protections possible through the use of enclaves, which are protected areas of execution in memory.




**Intel® Software Guard Extensions**

This Intel technology is for application developers who are seeking to protect select code and data from disclosure or modification. Intel® SGX makes such protections possible through the use of enclaves, which are protected areas of execution in memory.



**Intel® Software Guard Extensions**

This Intel technology is for application developers who are seeking to protect select code and data from disclosure or modification. Intel® SGX makes such protections possible through the use of enclaves, which are protected areas of execution in memory.



**Intel® Software Guard Extensions**

# 组内沟通情况

截止目前为止，小组组内的沟通畅通，协调较为统一，平时在阶段性工程完成之后，小组内会进行互相的说明通告，关于小组协作情况，需要说明的还有：

- 本周四，小组内对各个组员负责的github上的remote分支进行了merge，处理了相关的conflict
- 本周五，小组进行了一次有助教参与的stand meeting
- 具体的，小组内各组员负责的分支没有变化，工作继续推进：
  - 马轲-BackEnd：负责工程内后端程序（包括对前端访问的处理、数据库的处理、数据的排序查找等算法实现）的编写；
  - 谢添翼-PageRank：负责包括原先的PageRank相关内容在内的，包括对于可视化数据图生成和算法实现以及前端模块实现的相关工作；
  - 朱锦昊-FrontEnd：负责总体的前端网页界面的编写，包括但不限于视觉内容的实现，功能模块的实现等，并且与后端接口进行互动；
  - 曹金坤-Data：负责对于工程需要的数据内容的获取和处理，包括但不限于对于网络内容的爬取收集、数据处理已经对于原始数据对数据库的添加等工作。

# 本周遇到的问题

本周的开发过程中体现了一些问题：

- 因为之前没有网站的部署经验，所以在对本地工程进行在线部署的过程中，在服务器的环境搭建、域名绑定的步骤上花费了大量的时间进行资料检索和学习；
- 因为IEEE Xplore的反爬虫和限制访问策略，在进行大规模的数据爬取的过程中，即使使用了多线程的策略，但还是经常出现包括访问被限制等各种意外的情况，限制了数据爬取的效率；
- 在对数据匹配检索的算法实现中，因为数据量庞大，所以在使用包括TF-IDF算法在内的检索匹配策略时，遭遇了性能上的较大瓶颈，需要进一步的进行优化；
- 因为网页客户端适配和浏览器的不同，在包括部分移动端设备在内的部分情况下，网页的显示出现了一些和预想不同的情况，需要进一步优化；
- 在使用的Vue.js版本中，对于Jquery的兼容出现了问题，导致前端开发遭遇了一些困难；